UNIVERSITI PUTRA MALAYSIA

MOLECULAR ANALYSIS OF THE *EMM* GENE OF GROUP A STREPTOCOCCUS STRAIN DI323

REBECCA ROBERT RANTTY

FSAS 2001 53

MOLECULAR ANALYSIS OF THE *EMM* GENE OF GROUP A STREPTOCOCCUS STRAIN D1323

By

REBECCA ROBERT RANTTY

Thesis Submitted in Fulfilment of the Requirement for the Degree of Master of Science in the Faculty of Science and Environmental Studies Universiti Putra Malaysia

June 2001

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Science

MOLECULAR ANALYSIS OF THE *EMM* GENE OF GROUP A STREPTOCOCCUS STRAIN D1323

By

REBECCA ROBERT RANTTY

June 2001

Chairman : Associate Professor Khatijah Mohd Yusoff, Ph.D.

Faculty : Science and Environmental Studies

The epidemiological studies and characterization of group A streptococci (GAS) are mainly based on serological M and T typing, but although T typing is useful it is not M specific. In addition, it is difficult to prepare the M antisera and the increasing number of new M types makes them nontypeable with the available reference sera. The M protein is a major virulence factor of GAS which is encoded by the *emm* gene. The 5' ends of this gene are highly heterogenous and encode for specificity of the M serotypes used for M typing. Therefore sequencing of the 5' ends of the *emm* gene is the choice alternative to the serological typing in the characterization of GAS when M antisera are not available.

The Malaysian GAS strain, D1323 shows unique serotype specificity based on the homology searches of the 5' end *emm* gene sequence. The *emm* gene of D1323 was amplified using 'all M' primers and cloned into pCR®2.1-TOPO® vector for its sequence determination as well as into pTrcHis2-TOPO® vector for its expression. Plasmids of positive clones in pCR®2.1-TOPO® were sequenced and

the positive clones in pTrcHis2-TOPO® were analysed for protein expression by SDS-PAGE and Western immunoblotting.

The complete deduced sequence of the *emm* gene of D1323 was shown to contain an open reading frame of 1416 nucleotides which encodes for 429 amino acid residues of the mature M protein. There are three copies of C repeats in the sequence. The cleavage site of a signal peptide was predicted to be located at amino acid residue 42. Conserved regions of the C-terminus which are shared among various M serotypes and that of the leader peptide were also determined based on multiple sequence alignment. The M protein of D1323 was predicted as M Class I protein based on the alignment of the C-terminus and phylogenetic analysis. The fusion M protein was successfully expressed in the *Escherichia coli* system and its size was determined.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia
sebagai memenuhi keperluan untuk ijazah Master Sains

ANALISIS MOLEKULAR GEN *EMM* DARI STREPTOKOKUS KUMPULAN A STRAIN D1323

Oleh

REBECCA ROBERT RANTTY

Jun 2001

Pengerusi : Profesor Madya Datin Khatijah Mohd Yusoff , Ph.D

Fakulti : Sains dan Pengajian Alam Sekitar

Pengajian epidemiologi dan pengkelasan untuk Streptokokus Kumpulan A (GAS) adalah berdasarkan terutamanya kepada ujian serologi menggunakan pengtaipan antigen protein T dan M. Walaupun pengtaipan antigen T lazim dan meluas digunakan ianya bukan spesifik terhadap antigen protein M, di mana protein M ini merupakan faktor kevirulenan utama yang terlibat dalam opsonisasi untuk bakteria ini. Antisera untuk protein M adalah begitu sukar untuk disediakan, tambahan pula bilangan serotaip M yang baru sedang meningkat dan tidak boleh ditaipkan dengan antisera M yang sedia ada. Protein M adalah dikodkan oleh gen *emm*. Penghujung 5' gen ini adalah lebih heterogenous dan menentukan spesifisiti serotaip M yang digunakan untuk pengtaipan M. Dengan itu, penentuan jujukan penghujung 5' gen *emm* ini merupakan satu alternatif untuk ujian serologi bagi pengtaipan M di mana pengkelasan GAS dapat dilakukan tanpa antisera M.

Strain D1323 dari Malaysia menunjukkan spesifisiti serotaip yang unik berdasarkan penentuan homologi dengan jujukan gen untuk penghujung 5' dari

strain lain. Gen *emm* dari D1323 telah diklonkan dalam vector pCR®2.1-TOPO® untuk penentuan jujukan DNAnya dan juga ke dalam vektor pTrcHis2-TOPO® untuk ekspresinya. Jujukan DNA dari plasmid klon positif dalam pCR®2.1-TOPO® telah ditentukan and klon positif dalam pTrcHis2-TOPO® ditentukan ekpresinya dengan menggunakan SDS-PAGE dan pemblotan Western.

Daripada penentuan jujukan DNA D1323, ia mempunyai 1416 pasangan nukleotida yang akan mengkodkan 429 residu asid amino untuk protein M yang matang.. Terdapat tiga salinan ulangan C dalam jujukan DNAnya. Tapak pemotongan bagi peptida isyarat diramalkan terletak pada residu asid amino yang ke 42. Terdapat satu bahagian bahagian terperlihara pada C-terminus jujukan asid amino apabila dibandingkan dengan pelbagai serotaip M dan satu peptida isyarat untuk protein M bagi D1323 telah berjaya ditentukan. Protein M dari D1323 juga didapati tergolong dalam Kelas protein M I berdasarkan dari analisis C-terminus dan juga pilogenetiknya. Protein M telah berjaya diekspresikan dalam sistem *Escherichia coli* dan saiznya juga dapat ditentukan.

ACKNOWLEDGEMENTS

O, Lord God Almighty and Jesus Christ, thank you.

I would like to express my appreciation and sincere gratitude to my supervisor, Associate Professor Datin Dr. Khatijah Mohd. Yusoff for her invaluable guidance, encouragement, support and advice throughout the lab work of this project and in the preparation of this thesis.

I am also much indebted and grateful to my co-supervisors, Associate Professor Dr. Abdul Manaf Ali and Professor Datin Dr. Farida Jamal for their most kind guidance and helpful support.

My appreciation also goes to all my labmates in Lab 143, housemates and friends. Thanks for all the help and support for these past two years.

I would like to acknowledge my beloved family, papa, mama, Moses, Ani, Imelda, and Ini for all the love, support and trust that you gave me. I would never be able to go through this without all of you.

Thank you very much....

I certify that an Examination Committee met on 26th June 2001 to conduct final examination of Rebecca Robert Rantty on her Master of Science thesis entitled "Molecular Analysis of the *emm* Gene of Group A Streptococcus Strain D1323" in accordance with Universiti Pertanian Malaysia (Higher Degree Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the candidate be awarded the relevant degree. Members of the Examination Committee are as follows:

Raha Abdul Rahim, Ph.D.
Lecturer
Faculty of Food Science and Biotechnology
Universiti Putra Malaysia
(Chairman)

Khatijah Mohd. Yusoff, Ph.D.
Associate Professor
Faculty of Science and Environmental Studies
Universiti Putra Malaysia
(Member)

Abdul Manaf Ali, Ph.D.
Associate Professor
Faculty of Food Science and Biotechnology
Universiti Putra Malaysia
(Member)

Farida Jamal, Ph.D.
Professor
Faculty of Medicine and Health Science
Universiti Putra Malaysia
(Member)

MOHD GHAZALI MOHAYIDIN, Ph.D.
Professor/Deputy Dean of Graduate School
Universiti Putra Malaysia

Date : 11 JUL 2001

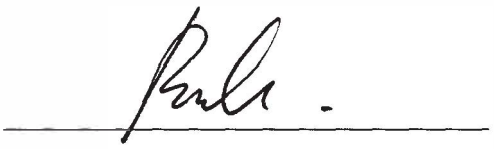
This thesis submitted to the Senate of Universiti Putra Malaysia has been accepted as fulfilment of the requirement for the Master Degree

AINI IDERIS, Ph.D.
Professor/
Dean of Graduate School
Universiti Putra Malaysia

Date :

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UPM or other institutions.



REBECCA ROBERT RANTTY

Date : 11th July 2001

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

ABBREVIATIONS

AP	-	Alkaline phosphatase
APS	-	Ammonium persulphate
BCIP	-	5-bromo-4-chloro-3-indolyl-phosphate disodium
BLAST	-	Basic Local Alignment Search Tool
bp	-	Base pairs
EDTA	-	Ethylenediaminetetraacetic acid
GAS	-	Group A streptococci
IPTG	-	Isopropyl-β-D-galactopyranoside
kb	-	Kilobase pairs
kDa	-	Kilodaltons
LB	-	Luria Bertani
MSA	-	Multiple sequence alignment
NBT	-	Nitro blue tetrazolium chloride
NCBI	-	National Center for Biotechnology Information
OD	-	Optical density
OF	-	Opacity factor
ORF	-	Open reading frame
PAGE	-	Polyacrylamide gel electrophoresis
PCR	-	Polymerase Chain Reaction
SDS	-	Sodium dodecyl sulphate
T,C,G,A	-	Thymine, Cytosine, Guanine and Adenine
TAE	-	Tris acetate EDTA buffer
TBE	-	Tris borate EDTA buffer
TEMED	-	N,N,N',N'-tetramethylethylenediamine
THB	-	Todd Hewitt broth
TTBS	-	Tris buffered saline with Tween
v/v	-	Volume over volume
w/v	-	Weight over volume
X-Gal	-	5-bromo-4-chloro-3-indolyl-3-D-galactoside

CHAPTER 1

INTRODUCTION

1.1 General Introduction

Streptococcus pyogenes of Lancefield Group A Streptococcus (GAS) is a Gram positive spherical bacterium which haemolyzes red blood cells, producing the characteristic beta-haemolytic colonies on blood agar. This highly successful pathogen is commonly found in the throat and on the skin. It is responsible for a variety of diseases in both adults and children, from a relatively mild sore throat to a more serious illness such as streptococcal toxic shock syndrome and necrotizing fasciitis (Fischetti, 1989). It is also noted for its non-suppurative sequelae, acute rheumatic fever and post-streptococcal glomerulonephritis (Bisno *et al.*, 1991).

In the late 1980s, a resurgence of severe GAS infections was noted in the Western Hemisphere (Cleary *et al.*, 1992; Musser *et al.*, 1993) although the increased incidence of GAS diseases remained undocumented in developing countries. This rekindled interest in GAS disease research worldwide.

The group A streptococcus cell is composed of an outer capsule, surface protein antigen, group specific carbohydrate, mucopeptide and cytoplasm. The outer capsule is composed of hyaluronic acid and the cell wall contains surface M, T and R proteins. The M protein is the major streptococcal virulence factor and

determines the serotype specificity of the isolates. The group specific carbohydrate is the one that defined the GAS from other streptococci. Hyaluronidase, DNase and streptokinase are all GAS surface products that help the organisms to adapt and spread through the tissues. These bacteria also produce extracellular products such as toxins and enzymes that might have the potential to act as virulence factor (Schmidt *et al.*, 1996).

For more than half of the last century, serotyping of the T and M surface antigens has been the standard method used for typing GAS – Lancefield, 1933 (Colman *et al.*, 1993; Seppala *et al.*, 1994). T agglutination is essential for initial screening as it is known that a certain particular T type implies the presence of certain M serotypes. However identification of T antigen in streptococcal epidemiology is not a substitute for the identification of the M antigen, primarily because the antibodies to the T antigen do not reflect type-specific immunity and also that the T antigen is often associated with multiple M types (Beall *et al.*, 1997). The opacity factor (OF) production can also be used to accomplish the characterization of GAS as it is consistently and exclusively associated with specific M serotype streptococcal strains (Johnson and Kaplan, 1988).

A large proportion of those isolates in the endemic areas such as in Malaysia are still M-nontypeable because of lack of reactivity towards available M antisera. Specific M typing sera are also difficult to obtain and very expensive to prepare. Only a small percentage of strains are typeable with standard M typing sera and

this suggests that the strains in this region belong to different and perhaps new M types (Tran *et al.*, 1994; Jamal *et al.*, 1995). Rapid identification and typing of isolates are essential for monitoring the spread and genetic variability of GAS. Therefore there is a need to develop an alternative means of M type deduction to the current serologic M typing (Beall *et al.*, 1996).

The M protein is encoded by the *emm* gene (Lancefield, 1962). The complete sequences of the *emm* and *emm*-like genes have been published (Robbins *et al.*, 1987; Mouw *et al.*, 1988; Haanes and Cleary, 1989). Even though the relationship between these genes vary in detail, they all possess a common framework; the 5' terminus of the *emm* gene comprises a highly conserved 5' leader peptide sequence with a hypervariable region of approximately 150 bp (Haanes and Cleary, 1989; Podbielski *et al.*, 1991; Whatman and Kehoe, 1994). This region encodes the peptides that protrude outwards from the cell surface. This sequence is followed by a highly conserved 3' region which seemed to be associated with the cell wall (Whatman and Kehoe, 1994). Each gene contains some internal repeats which vary in the extent and degree of the repetitions among different GAS strains. The peptides in the hypervariable region appear to be involved in the resistance towards phagocytosis and it is this diversity that forms the basis for M serotyping.

The M antibody corresponds to a specific M antigen but not to the other M types (Maxted and Valkenburg, 1968). Since the 5' termini are highly heterogenous and

are serospecific for a particular M protein, identification of such sequences may be used as an alternative to M serotyping.

The M protein is also the primary focus for vaccine development, but attempts to use this protein have been complicated by the extensive variability of the M antigen. Therefore it is important to identify regions of M protein molecule shared among various serotypes which might have potential for vaccine production (Mouw *et al.*, 1988). There are six provisional new *emm* types from Malaysia that suggest unique serologic specificity (Jamal *et al.*, 1999). One of these, D1323, is used in the study. It has less than 82% homology identity of the 5' termini to M protein of known sequence.

1.2 Objective

The objective of this study is to perform 5' *emm* sequence analysis (*emm* typing) on D1323, identify the regions which are shared among various M serotypes and express the M protein in *Escherichia coli*.

CHAPTER 2

LITERATURE REVIEW

2.1 Clinical Importance of GAS

Streptococcus pyogenes or Lancefield group A streptococci (GAS) is a versatile human pathogen, encountered worldwide which is responsible for a wide variety of infection both in children and young adults. Throat infection 'strep throat' (pharyngitis) which is relatively mild and if accompanied by a typical rash it is known as scarlet fever, is the most common infection caused by GAS. These infections can be followed by a more serious non-suppurative sequelae. One is acute rheumatic fever (ARF), a disease which primarily affects the heart and the other is post streptococcal glomerulonephritis, a serious condition in which the kidneys lose their ability to function properly. Another primary skin infection ie impetigo (pyoderma) which is most frequent, especially in tropical climates. Some of the rare but severe cases of necrotizing fasciitis caused by this bacteria have been linked with skin infections (Schmidt *et al.*, 1996).

A decrease in the incidence of mortality from rheumatic heart disease has been noted in Peninsular Malaysia (Khoo *et al.*, 1991); most probably a result of improved standards of living and better health facilities. However, in terms of human morbidity and mortality worldwide, the role of GAS in subsequent development of diseases is very important and continued surveillance of

rheumatic fever is indicated in view of its recent resurgence all over the world (Kaplan *et al.*, 1989; Martin and Hoiby *et al.,* 1990; Bisno *et al.*, 1991; Bouvet *et al.*, 1994). It is noted that the M types of GAS isolates prevalent in some ASEAN countries were different from those implicated in the 1980s resurgence in the Western Hemisphere (Kaplan *et al.*, 1992; Relf *et al.*, 1992).

Recent epidemiological findings (Cleary *et al.*, 1992; Musser *et al.,* 1993) have reinforced the interest in this species and led to reassessment of the efficiency and significance of the methods for characterization. It is very important to monitor GAS strains especially in this geographical area as accurate identification and characterization of GAS is useful in the study of its epidemiology, pathogenicity and therapy of infection.

2.2 Structure and Antigenic Composition of GAS

2.2.1 Cell Structure

The cell structure of GAS (Figure 2.1) includes several important components. The outermost layer is the capsule consisting of hyaluronic acid and it forms a slimy outercoat that retards phagocytosis by leukocytes. The cell wall surface, a layer in the absence of the capsule is covered with hair-like protrusions or fimbriae contain the M, T and R antigens and lipoteichoic acid (Krause, 1972). M protein makes up 30-50% of the dry weight of the cell. Lipoteichoic acid is a

polymer consisting of repeating units of glycerophosphate and a terminal glycolipid. It facilitates the adherence of GAS to the mucous membrane of the cell surface such as pharyngeal epithelium (Fischetti, 1989). The biological functions of R and T protein antigens in virulence are still unknown although they are very useful epidemiological markers (Haanes and Cleary, 1989).

The cell wall consists of rhamnose-*N*-acetyl glucosamine polysaccharide. This group specific carbohydrate forms the basis of serologic grouping (Lancefield, 1933). Beneath the cell wall is the cytoplasm but the antigens of the cytoplasmic membrane are not utilized in the classification of GAS.

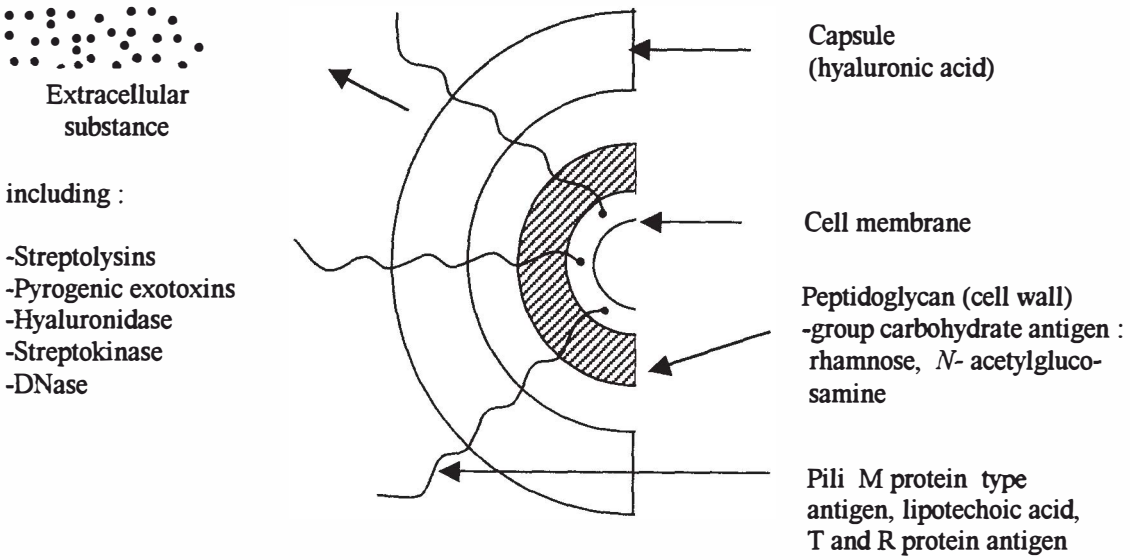


Figure 2.1: Diagrammatic representation of the subcellular component of GAS (Krause, 1972)

2.2.2 Antigenic and Virulence Factors of GAS

GAS express a range of cell surface and extracellular products which have the potential to act as virulence factors contributing to pathogenicity. The components protecting these bacteria against phagocytic attack play an important role in pathogenesis.

The streptococcal M protein antigen is considered to be one of the major surface components responsible for the resistance of GAS against phagocytosis (Lancefield, 1962). It is also involved in adhesion (Fischetti, 1989). The hyaluronic acid capsule in streptococci forms a slimy outer coat that retards phagocytosis by leukocytes. This capsule is non-antigenic but it helps in resisting the opsonizing antibodies (Schmidt *et al.*, 1996). There is no evidence that the other two surface products, the T and R antigens, are involved in virulence.

The pathogenic streptococci also produce a number of toxins and enzymes as their extracellular products (Figure 2.1). Antigenic streptococcal exotoxins secreted by GAS are erythrogenic toxins A, B and C which are responsible for the scarlet fever rash. Streptolysin-S is a non-antigenic toxin which produces haemolysis around the colonies while Streptolysin-O is a reversibly oxygen-labile cytolysin with cardiotoxic potential and it is not active in the presence of O_2. Hyaluronidase, DNase and streptokinase are all GAS enzymes that help these organisms to adapt and spread through tissues (Ashbaugh *et al.*, 1998).

Streptokinase activates the plasma enzyme (plasminogen) to become an active protease (plasmin) which digests fibrin clots. Hyaluronidase, digest the ground substance of connective tissue and aids the movement of organisms through tissues. DNase hydrolyses and thin out the viscous deposits of DNA. Infections due to GAS often result in thin, spreading exudates rather than a thick pus of the well-localized abscesses.

2.3 Serological Typing for Serotype Characterization

2.3.1 OF Detection and OF Inhibition Typing

Some strains belonging to certain M types produce opacity in mammalian sera as a result of the action of an apoproteinase, an enzyme referred to as serum opacity factor (OF) on a high-density lipoproteins (Maxted *et al.,* 1973). The OF of each M type of the GAS (Top and Wannamaker, 1968; Hallas and Widdowson, 1983) is constant and has the same antigenic specificity. In other words, a particular M-type not only produces a specific OF type, it also induces type-specific OF antibodies that can be used for OF inhibition typing. Approximately half of the known M serotypes express OF (Hill and Wannamaker, 1968). GAS can be categorized into two broad groups, OF positive and OF negative, based on the presence or absence of the serum OF (Hannes *et al.,* 1992).